5/26/17

"What goes up doesnt always have to come down." -Unknown

HW: "Intro to Statistics" homework section #1-6

January 2017 #1-24

Test 3 on Wednesday 6/7

AIM: What are Statistics?

Warm Up:

opposite

1. The graph of $y = 3\sin(3x - \frac{\pi}{3}) - 4$ is a transformation of $y = 3\sin(3x)$. Describe the transformation completely.

shift right $\frac{\pi}{3}$ units

down 4 units

APPLICATIONS

1. A fair coin is flipped four times. Find:

(a) The probability it will land up heads each time.

$$P(\text{H}_1 \text{ and } \text{H}_2 \text{ and } \text{H}_3 \text{ and } \text{H}_4) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{16}$$

(b) The probability it will land the same way each time (slightly different from (a)).

$$P((\text{H}_1 \text{ and } \text{H}_2 \text{ and } \text{H}_3 \text{ and } \text{H}_4) \text{ or } (\text{T}_1 \text{ and } \text{T}_2 \text{ and } \text{T}_3 \text{ and } \text{T}_4)) = \frac{1}{16}+\frac{1}{16} = \frac{2}{16} = \frac{1}{8}$$

2. A first grade class of nine girls and seven boys walks into class in alphabetical order (by last name). What is the probability that three girls are the first to enter the room? Show your calculation.

(1) 0.15 (3) 0.35

(2) 0.20 (4) 0.45

$$\frac{9}{16}\cdot\frac{8}{15}\cdot\frac{7}{14} = \frac{504}{3360} = 0.15$$

(1)

3. A bag of marbles contains 12 red marbles, 8 blue marbles, and 5 green marbles. If three marbles are pulled out, find each of the following probabilities. In each we specify either replacement (the marbles go back into the bag after each pull) or no replacement.

(a) Find the probability of pulling three green marbles out with replacement.

$$\frac{5}{25}\cdot\frac{5}{25}\cdot\frac{5}{25} = \frac{125}{15{,}625} = \frac{1}{125}$$

(b) Find the probability of pulling out 3 red marbles without replacement.

$$\frac{12}{25}\cdot\frac{11}{24}\cdot\frac{10}{23} = \frac{1{,}320}{13{,}800} = \frac{11}{115}$$

(c) Find the probability of pulling out 3 marbles of the same color without replacement. This is more complex than the other two.

$$P(\text{all red or all blue or all green})$$

$$P(\text{all red}) = \frac{1{,}320}{13{,}800}$$

$$P(\text{all blue}) = \frac{8}{25}\cdot\frac{7}{24}\cdot\frac{6}{23} = \frac{336}{13{,}800}$$

$$P(\text{all green}) = \frac{5}{25}\cdot\frac{4}{24}\cdot\frac{3}{23} = \frac{60}{13{,}800}$$

$$P(\text{all red or all blue or all green}) = \frac{1{,}320}{13{,}800}+\frac{336}{13{,}800}+\frac{60}{13{,}800} = \frac{1{,}716}{13{,}800}$$

(d) Find the probability of pulling out two blue marbles and one green marble in any order with replacement. Be careful as there are multiple ways this can be done that will add.

$$P(\text{BBG or BGB or GBB}) = \frac{8}{25}\cdot\frac{8}{25}\cdot\frac{5}{25}+\frac{8}{25}\cdot\frac{5}{25}\cdot\frac{8}{25}+\frac{5}{25}\cdot\frac{8}{25}\cdot\frac{8}{25}$$
$$= \frac{320}{15{,}625}+\frac{320}{15{,}625}+\frac{320}{15{,}625} = \frac{960}{15{,}625} \approx 0.06$$

4. The table below shows the percents of graduating seniors who are going to college, broken down into subgroups by gender. If a student was picked at random find the probability that:

(a) They would be a female going to college.

	Percent of Graduating Seniors	Percent of Subgroup Going to College
Male	46%	78%
Female	54%	84%

$$\begin{aligned} P(\text{F and C}) &= P(\text{F})\cdot P(\text{C}|\text{F}) \\ &= (0.54)(0.84) \\ &= 0.4536 \end{aligned}$$

(b) They would be a male not going to college.

$$\begin{aligned} P(\text{M and not C}) &= P(\text{M})\cdot P(\text{not C}|\text{M}) \\ &= (0.46)(0.22) \\ &= 0.1012 \end{aligned}$$

(c) They would be going to college.

$$\begin{aligned} P((\text{M and C}) \text{ or } (\text{F and C})) &= P(\text{M})\cdot P(\text{C}|\text{M}) + P(\text{F})\cdot P(\text{C}|\text{F}) \\ &= (0.46)(0.78) + (0.54)(0.84) \\ &= 0.3588 + 0.4536 \\ &= 0.8124 \end{aligned}$$

(d) They would not be going to college.

$$\begin{aligned} P(\text{not C}) &= 1 - P(C) \\ &= 1 - 0.8124 = 0.1876 \end{aligned}$$

5. If a safety switch has a 1 in 10 chance of failing, how many switches would a company want to install in order to have only a 1 in one million chance of them all failing at the same time? Show your reasoning.

Let the number of switches be n. Then we could set up the following probability that all n-switches fail:

$$P(\text{S}_1 \text{ and S}_2 \text{ and ... S}_n) = \frac{1}{10}\cdot\frac{1}{10}\cdot\ldots\frac{1}{10} = (0.1)^n$$

$$\left(\frac{1}{10}\right)^n = \frac{1}{1,000,000} \Rightarrow 10^n = 1,000,000 \Rightarrow n = 6$$

Reasoning

6. If the probability of winning a carnival game was $\frac{2}{5}$ and Max played it five times, write an expression that would calculate the probability he won the first three games and lost the last two. Use exponents to express your final answer, but do not evaluate.

$$\begin{aligned} P(\text{w}_1 \text{ and w}_2 \text{ and w}_3 \text{ and l}_4 \text{ and l}_5) &= \frac{2}{5}\cdot\frac{2}{5}\cdot\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{3}{5} \\ &= \left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^2 \end{aligned}$$

Data is everywhere. It's in our newspapers, it's in our science classes, it shows up in economics, medicine and anywhere else that **variability** occurs. Variability is simply the property of **outcomes** being different. The tools of statistics are designed to **explain this variability**.

There are many types of variability. It is good to understand these sources in order to minimize the ones that we are not studying.

(a) **Observational or Measurement Variability**: Variability that is introduced due to either our measuring instruments not being precise enough or differences in how two different people read the measurement.

reaction time of 2 people timing a race.

(b) **Natural Variability or Inter-Individual Variability**: Variability that accounts for the fact that members of a populations are simply different.

Two people eat the same food but only one gains weight.

NOT NATURAL

(c) **Induced Variability**: This type of variability is in marked contrast to natural. It occurs because we have assigned our population or sample to two or more **treatment** groups and then observe the variability between the groups. (making the groups different)

Testing memory of two groups – one group is allowed to sleep, the other not, + then they are tested.

(d) **Sample Variability**: This is the type of variability that occurs when we take multiple **samples** from a **population** randomly. These samples will be different due to the randomness of the sampling process.

Population = everybody
sample population - smaller group (sample) from population

Population of U.S. – everyone
sample ex: Nursing home residents
or math class students
OR NY Democrats & Repub.
V.S. MONTANA " "

Remember, through all of our work in this unit, we are really trying to explain the variability of data within either a population or a sample and then using this to determine if the variability can be attributed to one of the factors above to the exclusion of the others.

There are many different situations in which we collect data. They have important differences and all of them depend on randomization in one way or another. *not knowing, random

The three major types of ways to collect data are described below. Let's give an example of each and explain how **randomization** is part of each method. Randomization is used primarily to eliminate variability caused by some type of bias. *something not fair

(a) **Surveys**: $\boxed{\text{Collections of data}}$ from a population where variability is not induced by treatments but by the sample itself (sampling variability). ASK Questions to a random sample

Poll → survey
phone questions = survey

Just watching - no asking

(b) **Observational Studies**: Collections of data from a population where assignment of individuals from the population into **treatment groups** is **not** under the control of those performing the study.

Look at data

(c) **Experimental Studies**: In experimental studies individuals are assigned randomly to treatment groups in order to determine the effect of the treatment on the variability of the data. In these cases, the assignment, although random, is under the control of those performing the study.

multiple groups - ① Control Group
② Test Group
} Subjects chosen randomly

Random sampling is critical for being able to minimize variability due to **sampling bias**. Random sampling can be done using a variety of different techniques. Simple random sampling can be accomplished using a random number table.

***Exercise* #1:** A list of 10 people's heights, in inches, is shown below.

Person #	1	2	3	4	5	6	7	8	9	10
Height	70	68	60	75	65	69	58	62	66	63

(a) Randomly select five heights from this list by using the random number table that goes with this lesson. Choose a random spot in the table and move down the column. Select the first digit of each number. If you get a repeat, eliminate and keep going. If you get a 0, use this as the 10.

68, 75, 69, 62, 63

(b) Calculate the **sample mean** to the nearest tenth. Compare to others in the class. What type of variability is being introduced through this process?

$$\frac{68+75+69+62+63}{5}$$

≈ 67.4

Ex: Sampling variability

When we conduct a study, the complete set of all subjects that share a common characteristic that is being studied is known as the **population**. All populations have **natural or inter-individual variability**. Most of the time, the entire population is not measured, but a sample is taken to infer characteristics of a population. Still, all populations in theory have **population parameters** that describe the population, such as its mean, standard deviation, and interquartile range.

***Exercise* #2:** 18 students in Mr. Weiler's Advanced Calculus class took a quiz with the following results in ascending order.

56, 68, 72, 72, 75, 78, 80, 84, 84, 85, 88, 88, 90, 93, 95, 99, 100, 100

(a) Use your calculator to determine the mean, the median, and the quartiles for this data set. Then, construct a simple box-and-whiskers (box plot) for this data set.

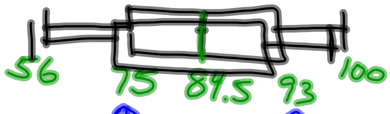
```
NORMAL FLOAT FRAC REAL RADIAN MP
1-Var Stats
x̄=83.72222222
Σx=1507
Σx²=128637
Sx=12.04797057
σx=11.7085228
n=18
minX=56
↓Q1=75
```

mean

```
NORMAL FLOAT FRAC REAL RADIAN MP
1-Var Stats
↑Sx=12.04797057
σx=11.7085228
n=18
minX=56
Q1=75
Med=84.5
Q3=93
maxX=100
```

1st Quartile 3rd

median



IQR

(b) What is the interquartile range of this data set? In theory, what percent of the data set should lie between the first and thi~~rd quartiles? Is that~~ true for this data set?

$IQR = Q_3 - Q_1$

$93 - 75 = \boxed{18}$

50%

$8/18 \rightarrow$ 44% for this

(c) What is the population standard deviation for this data set to the nearest tenth? How do you interpret the standard deviation?

$\sigma x = 11.7$

On average a data point is 11.7 units from the mean.

(d) What percent of the scores were within one standard deviation of the mean? Within two standard deviations of the mean? Round your percents to the nearest percent and show your work.

Within One Standard Deviation of the Mean

$\bar{x} + \sigma x$

$83.7 + 11.7 = 95.4$

$\bar{x} - \sigma x =$

$83.7 - 11.7 = 72$

11 out of 18

$\frac{11}{18}$ or 61%

Within Two Standard Deviations of the Mean

$8\overline{3.7} + 2(11.7) = 107.1$

$83.7 - 2(11.7) = 60.3$

17 out of 18

$\frac{17}{18}$ or 94%

Sometimes data is grouped in a frequency chart. We still should be able to calculate the basic population parameters when the information is given in this form.

***Exercise* #3:** A small company has salaries for their 50 employees as given in the table below

Salary (x_i) L_1	Frequency (f_i) L_2
25,000	5
32,000	21
45,000	14
58,000	7
75,000	2
120,000	1

(a) Find the mean and standard deviations of the salary range.

$\bar{x} = 42060$ (mean) $\sigma x = 16567.93$ (std deviation)

(b) What is the median of this data set? Why is the median considerably lower than the mean in this data set?

median = 32000

Because there are 21 $32 000 employees

(c) Does more or less than 50% of the data set fall within one standard deviation of the mean? Show the analysis that leads to your answer.

$\bar{x} + \sigma x = 58627.93$

$\bar{x} - \sigma x = 25492.07$

42 out of 50

84%

Although we have often concentrated on experimental studies where data is collected and means are found, many times we use statistics to represent results of a survey where we are interested in what **proportion** of a **population** share a certain characteristic. These proportions are most expressed as decimals, but sometimes are represented by fractions or percents.

***Exercise* #4:** A questionnaire went home to all juniors concerning their ability to bring and use mobile devices at school. The questionnaires constituted a **census** since all of the juniors were surveyed. Of the 742 juniors, 564 of them reported having web-enabled mobile devices. What was the population proportion for web-enabled devices? Express your answer as a decimal and as a percent.

***Exercise* #5:** The proportion of eggs that get cracked in a local egg handling facility is 0.023. If 2,500 dozen eggs are packaged in the factor per day, what should we expect to be the number of eggs cracked per day?

(1) 350 (3) 230

(2) 450 (4) 690